Module: Theory and Practice of Modern Distributed Systems

The course covers the foundational theory of distributed systems, and extends to the modern practice of distributed systems used to design NoSQL databases, Big Data platforms, blockchain technologies and federated machine learning.

The course covers a wide spectrum of topics such as clocks and synchronization, programming, consensus and blockchains, mutual exclusion, distributed databases and commit protocols. The course will be spread around five modules as listed below with each module having four to five hours of lectures and three to four hours of practicum. There will be an emphasis on the practical use of contemporary distributed systems software such as distributed storage, Big Data volume and velocity processing platforms, and distributed ML platforms.

The fifth module has two variants: one variant that connects distributed computing to ML, and the other variant connects distributed computing to ideas such as blockchain. Participants can choose to attend one of the two variants and these two variants happen in parallel. The table below lists the details of the lecture and practice component of each module. K refers to Kishore and Y refers to Yogesh, in terms of the lecture, practice, and overall coordination.

The entire course will be run in-person from IIIT Hyderabad during July 3 to 15, 2022. Participants will be provided options for staying in campus during that time.

S. No.	**Module Name**	**Lecture component (L)**	**Practicum Component (P)**	**Duration (Hrs)**	
				L	P
1	Principles of Distributed Systems	Time, Synchronization, Mutual Exclusion?, Communication	MPI, NTP, Truetime	8	4
2	Distributed File Systems and variants	NFS, GFS, CEPH, Haystack, Tectonic etc.	Using NFS, GFS, and other distributed file systems	8	4
3	Distributed Databases	Distributed databases,	Bigtable, Cassandra, 2PC with fault simulation	8	8

		ACID vs BASE, 2PC, NOSQL, Bigtable, Dynamo, CAP Theorem, Cassandra			
4	Big Data Systems and Processing	Data volume using Spark, Publish-Subscribe using Kafka, Data stream processing and Spark streaming. Applications from IoT, Streaming algorithms	Spark Data Frames/MapReduce, Kafka, Spark Streaming	8	8
5A	ML and Distributed Systems.	Scaling ML vs. DNN. Data and model parallelism. Parameter server. Federated learning, GNN	Spark ML, FedML, distributed ML using PyTorch	8	8
5B	Distributed Consensus and Blockchain	Achieving Consensus, RAFT. Distributed ledger and blockchain. Smart contracts.	Dragonboat Hyperledger Fabric	8	8